Large Deviation Principle for Occupancy Problems with Colored Balls

Paul Dupuis*
Lefschetz Center for Dynamical Systems
Brown University
Providence, R.I. 02912
USA

Carl Nuzman
Bell Labs
600 Mountain Avenue
Murray Hill, New Jersey 07974
USA

Phil Whiting
Bell Labs
600 Mountain Avenue
Murray Hill, New Jersey 07974
USA

June 6, 2003

Abstract

A Large Deviations Principle (LDP), demonstrated for occupancy problems with indistinguishable balls, is generalized to the case in which balls may be distinguished by a finite number of colors. The colors of the balls are chosen independently from the occupancy process itself. There are r balls thrown into n urns with the probability of a ball entering a given urn being $1/n$ (Maxwell-Boltzman statistics). The LDP applies with the scale parameter n going to infinity and the number of balls increasing proportionally. It holds under mild restrictions, the key one being that the coloring process by itself satisfies a LDP. Hence the results include the important special cases of deterministic coloring patterns and of colors chosen with fixed probabilities independently for each ball.

1 Introduction

In occupancy models which follow Maxwell-Boltzman statistics, balls are thrown in to n urns with the probability of a ball entering a given urn being

*This research was supported in part by the National Science Foundation (NSF-DMS-0072004, NSF-ECS-9979250) and the Army Research Office (DAAD19-02-1-0425).

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **06 JUN 2003**	2. REPORT TYPE	3. DATES COVERED **00-06-2003 to 00-06-2003**

4. TITLE AND SUBTITLE **Large Deviation Principle for Occupancy Problems with Colored Balls**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Brown University,Division of Applied Mathematics,182 George Street,Providence,RI,02912**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**32**	

$1/n$, independently of all other balls. References [1] and [4] develop sample path large deviations principles for scaled occupancy processes in which the time variable is (approximately) the number of balls thrown per urn, the state is given by the fraction of urns which contain exactly i balls, and the number of balls and urns are scaled up in fixed proportion. An LDP is obtained for infinite-dimensional processes in [1], whilst [4] focuses on processes with a finite number of occupancy levels, $i = 0, \ldots, I$ with $i = I+$ for urns with more than I balls. Additionally [4] provides explicit solutions to the corresponding calculus of variations problem.

In this paper we consider a generalization of such occupancy models to allow balls with more than one color. We fix on the case of two colors, as the extension to any finite number of colors is straightforward. The overall process can be regarded as the conjunction of two independent random processes, an occupancy process which determines which urn each ball enters, and a second process that determines color. The coloring process can be quite general and includes the important special case where each color is picked independently and according to a fixed vector of probabilities (iid coloring), as well as deterministic coloring patterns. Again time is scaled by a factor of n, so that at time $t \approx k/n \in [0, \beta]$, k balls have been thrown. The state of the process is the empirical measure, which records the fraction of urns that contain i balls of color 1 and j balls of color 2, for $0 \leq i \leq I+$ and $0 \leq j \leq J+$, where $I+$ and $J+$ correspond to more than I balls and more than J balls, respectively. Thus $\Gamma^n_{i,j}(t)$ is the fraction of urns containing i color 1 balls and j color 2 balls after approximately nt balls have been thrown. In general the process $\{\Gamma^n_{i,j}(t)\}$ will not be Markov unless the coloring process itself is Markov.

There is a wide literature on occupancy problems, and the case of distinguished classes of balls is a common generalization [7, 8]. A recent motivating application for colored occupancy problems is the analysis of wavelength conversion in the optical packet switch described in [6]. In each time slot, a random collection of packets (balls) arrive on a set of input fibers and must be routed onto a set of output fibers (urns). The packets on each fiber are wavelength-multiplexed on a finite number of channels (colors). In the absence of wavelength conversion, packets must use the same channel on the input and output fiber. If multiple packets of the same channel belong to the same output fiber, the excess packets must be converted to a different channel or discarded. Typical quantities to be computed include the probability of requiring a large number of wavelength converters and the probability of discarding a large number of packets. The problem was approached with single color large deviation analysis in [6]; the results there give only an

upper bound on the true number of converters because packets discarded due to fiber capacity were also considered to require conversion. By contrast, the multi-colored analysis contains the information needed to avoid this overestimation by taking fiber capacity into account. The multi-colored approach may also be useful in studying packet switches with constrained wavelength conversion patterns.

An example in statistics is the *problem of coincidence*, which may be illustrated by the following example, see also [7]. Let days in a given period be urns, and let there be three colors indicating the asthma attacks (balls) of three individuals. A ball goes into an urn if the individual has an asthma attack on that day. If there is no common cause (e.g., pollution event) the distribution of attacks are independent, and we are interested in the probability that the actual distribution could have arisen.

We derive the LDP for the colored occupancy processes by using the representation theorem for the scaled log moment generating functions for measurable functions of sample paths, see [3]. The representation is as an infimum over measures of the sum of a relative entropy cost and a terminal cost. As discussed in Section 2, there is a natural split of the relative entropy cost between a cost for occupancy and one for the coloring process. The local rate function in the single-color case is the relative entropy $R\left(\Theta(t) \mid\mid \Gamma(t)\right)$, where $\Theta_i(t)$ is the rate at which balls enter i-occupied urns when the time is t [4]. The corresponding expression in the colored case is a weighted sum of relative entropy terms $\dot{x}_1 R\left(\Theta^1(t) \mid\mid \Gamma(t)\right) + \dot{x}_2 R\left(\Theta^2(t) \mid\mid \Gamma(t)\right)$, where $\Theta^k_{i,j}(t)$ is the normalized rate at which balls of color k enter urns that presently contain i balls of color 1 and j balls of color 2, and where $x_i(t)$ is the fraction of color i balls per urn by time t. The overall local rate function also includes an additional term not present in the single color case, namely the local rate function for the coloring process itself.

In [4], the large deviations upper bound followed from the results in [2], but in the present case this is no longer true since the occupancy process need not be Markov. Instead, we present a direct proof based on weak convergence which only assumes a sample path LDP for the coloration process.

The most significant obstacle to obtaining an LDP occurs in the proof of the large deviations lower bound. The difficulty here is the singular behavior of the relative entropy cost when any element of $\Gamma(t)$ approaches zero. In [4], this difficulty is met in two steps. The boundary is avoided everywhere, except at the initial point, using a perturbation argument which relies on the joint convexity of the local rate function. A simple "filling" construction is then employed in the vicinity of the initial point. The construction is essentially equivalent to the construction of a change of measure with properly

bounded Radon-Nikodym derivative that would be needed in the traditional approach to the large deviation lower bound. In the present setting, a more delicate perturbation argument is required because the local rate function is not always jointly convex, as may be verified with simple examples (e.g. coloring via a two-state Markov chain). In addition, the filling construction is replaced by one based on time-reversal.

A striking feature of the single color occupancy analysis is that the calculus of variations problem can be solved to obtain explicit extremal trajectories, and the rate function for the final occupancy state can be computed directly by solving a fixed-point equation [4]. The same is true in the present case, and a companion paper [5] is in preparation that generalizes the calculus of variations analysis to colored balls.

An outline of the paper is as follows. In Section 2 a precise formulation of the model is given and the main results of the paper - the upper and lower Laplace principles - are stated. These bounds are equivalent to the large deviation upper and lower bounds. The section also presents three important special cases for coloration processes. In Section 3 the upper bound is established and Section 4 establishes some properties of the rate function which will be needed in the proof of the lower bound in Section 5.

2 Preliminaries and Main Result

We construct an urn model with colored balls as follows. Balls are thrown into one of n urns sequentially. The throwing process is modeled by a collection of independent and identically distributed (iid) random variables $\{X_l^n, l = 1, \ldots, \lfloor n\tau \rfloor + 1\}$, where $\lfloor a \rfloor$ denotes the integer part of the scalar a. Each X_l^n is uniformly distributed on the set $\{1, \ldots, n\}$, with each value of the set corresponding to an urn. Thus a total of $N_n \doteq \lfloor n\tau \rfloor + 1$ balls are thrown. There is also a coloration process designated by $Y_l^n \in \{1, 2\}$. At each discrete time a ball is assigned color Y_l^n, and then placed into urn number X_l^n.

We form empirical measures $\Gamma_{i,j}^n(t)$ as follows. If $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J\}$, then

$$\Gamma_{i,j}^n(l/n) \doteq \frac{1}{n} \left(\sum_{m=1}^{n} 1_{\left\{\sum_{r=1}^{l} 1_{\{X_r^n = m, Y_r^n = 1\}} = i\right\}} 1_{\left\{\sum_{r=1}^{l} 1_{\{X_r^n = m, Y_r^n = 2\}} = j\right\}} \right).$$

In other words, $\Gamma_{i,j}^n(l/n)$ is the fraction of cells containing exactly i color 1 and j color 2 balls when l balls have been thrown. Similarly,

$$\begin{aligned}
\Gamma^n_{I+,j}(l/n) &\doteq \frac{1}{n}\left(\sum_{m=1}^n 1_{\left\{\sum_{r=1}^l 1_{\{X^n_r=m,Y^n_r=1\}}>I\right\}} 1_{\left\{\sum_{r=1}^l 1_{\{X^n_r=m,Y^n_r=2\}}=j\right\}}\right)\\
\Gamma^n_{i,J+}(l/n) &\doteq \frac{1}{n}\left(\sum_{m=1}^n 1_{\left\{\sum_{r=1}^l 1_{\{X^n_r=m,Y^n_r=1\}}=i\right\}} 1_{\left\{\sum_{r=1}^l 1_{\{X^n_r=m,Y^n_r=2\}}>J\right\}}\right)\\
\Gamma^n_{I+,J+}(l/n) &\doteq \frac{1}{n}\left(\sum_{m=1}^n 1_{\left\{\sum_{r=1}^l 1_{\{X^n_r=m,Y^n_r=1\}}>I\right\}} 1_{\left\{\sum_{r=1}^l 1_{\{X^n_r=m,Y^n_r=2\}}>J\right\}}\right).
\end{aligned}$$

By definition $\Gamma^n_{0,0}(0) \doteq 1$, and $\Gamma^n_{i,j}(0) \doteq 0$ for all other values of (i,j). (One can also consider other initial conditions, with only simple notational changes in the results to be stated below. When extended to accommodate general initial conditions, the large deviation results we will prove are uniform in the initial condition, in the sense used in [4].) The definition of Γ^n is extended to all $t \in [0,\tau]$ not of the form l/n by piecewise linear interpolation. Let $\mathcal{U}$ denote the set of all probability measures on $\{0,1,\ldots,I,I+\} \times \{0,1,\ldots,J,J+\}$. The processes Γ^n are considered to take values in the space of continuous functions $\mathcal{S} \doteq C([0,\tau]:\mathcal{U})$, equipped with the usual supremum norm.

We wish to analyze the large deviation asymptotics of these processes, when the underlying coloration process satisfies a large deviation principle and is independent of the urn selection. To this end, it is convenient to use the Laplace formulation. Let F be any bounded and continuous function on $\mathcal{S}$. The processes Γ^n are said to satisfy a Laplace principle with rate function I if the following two conditions hold:

- For each $M < \infty$, the set $\{\Gamma : I(\Gamma) \le M\}$ is compact in $\mathcal{S}$.
- $$\lim_{n\to\infty} -\frac{1}{n} \log E \exp\left[-nF(\Gamma^n)\right] = \inf_{\Gamma\in\mathcal{S}} \left[I(\Gamma) + F(\Gamma)\right].$$

Since the processes Γ^n take values in a Polish space, the notions of Laplace principle and large deviation principle are equivalent [3, Corollary 1.2.5].

Cumulative coloration processes $\{x^n, n \in \mathbb{N}\}$ are defined for $t = l/n$ by

$$x_1^n(l/n) \doteq \frac{1}{n}\sum_{r=1}^l 1_{\{Y^n_r=1\}}, \quad x_2^n(l/n) \doteq \frac{1}{n}\sum_{r=1}^l 1_{\{Y^n_r=2\}}.$$

These definitions are also extended to $t \in [0,\tau]$ not of the form l/n by piecewise linear interpolation. Define the set of functions $\mathcal{T}$ by $x = (x_1, x_2) \in \mathcal{T}$ if

- $x_k(\cdot)$ is increasing and continuous with $x_k(0) = 0$ for $k = 1, 2$,
- $x_1(t) + x_2(t) = t$ for all $t \in [0, \tau]$.

We consider this set of functions as endowed with the usual sup norm topology, and make the following assumption.

Assumption 2.1 *The sequence of coloration processes $\{x^n, n \in \mathbb{N}\}$ satisfy a large deviation principle on $\mathcal{T}$ with the rate function J.*

Since $\mathcal{T}$ is also a Polish space, as noted previously this is equivalent to the statement that J satisfies the corresponding Laplace principle: J has compact level sets, and for all bounded and continuous functions $G : \mathcal{T} \to \mathbb{R}$,

$$\lim_{n\to\infty} -\frac{1}{n} \log E \exp\left[-nG(x^n)\right] = \inf_{x\in\mathcal{T}} \left[J(x) + G(x)\right].$$

Additional assumptions on the coloration processes will be introduced below. In particular, a mild structural assumption on J must be assumed in order to prove the large deviation lower bound.

We next describe a few typical coloration processes. The relative entropy function will be used for this purpose, and indeed throughout the paper. For two probability measures α and β on a Polish space $\mathcal{A}$, the relative entropy of α with respect to β is defined by

$$R\left(\alpha\|\beta\right) \doteq \int_{\mathcal{A}} \frac{d\alpha}{d\beta} \left(\log \frac{d\alpha}{d\beta}\right) d\beta = \int_{\mathcal{A}} \left(\log \frac{d\alpha}{d\beta}\right) d\alpha$$

whenever α is absolutely continuous with respect to β (and with the convention that $0 \log 0 = 0$). In all other cases we set $R\left(\alpha\|\beta\right) = \infty$.

Example 2.1 *Suppose we color the balls to achieve a deterministic fraction p_k of color k, with $p_k \in (0, 1)$. More precisely, if N_{l-1}^k balls of color k have been thrown in the first $l - 1$ throws (with $N_{l-1}^1 + N_{l-1}^2 = l - 1$), and if $N_{l-1}^1/n \leq p_1 l/n$, then we color the lth ball 1, and otherwise color it 2. The rate function for the corresponding processes $\{x^n, n \in \mathbb{N}\}$ is quite simple:*

$$J(x) \doteq \begin{cases} 0 & \text{if } x_k(t) = p_k t \\ \infty & \text{else.} \end{cases}$$

Example 2.2 *An alternative coloring scheme is to select the color in an iid fashion, with probability p_k of color k, where $p_k \in (0, 1)$. If a is a probability vector define*

$$M(a) \doteq R(a\|p),$$

and in all other cases let $M(a) \doteq \infty$. Then the rate function is

$$J(x) \doteq \begin{cases} \int_0^\tau M(\dot{x}_1(t), \dot{x}_2(t))dt & \textit{if } x \textit{ is absolutely continuous} \\ \infty & \textit{else.} \end{cases}$$

Example 2.3 *In our final example the color is determined by a two-state ergodic Markov process. Let the underlying transition probabilities be denoted $p_{k,l}, k = 1,2, l = 1,2$. Let b denote the invariant distribution, with $b_k \in (0,1)$. Given a probability vector a, let $q_{k,l}$ be any ergodic probability transition matrix with invariant distribution a. Define*

$$M(a) \doteq \inf \sum_{k=1}^{2} R\left(q_{k,\cdot} \| p_{k,\cdot}\right) a_k,$$

where the infimum is over all such transition matrices q. In all other cases set $M(a) \doteq \infty$. Note that $M(a) = 0$ if and only if $a = b$. Here again the rate function is written

$$J(x) \doteq \begin{cases} \int_0^\tau M(\dot{x}_1(t), \dot{x}_2(t))dt & \textit{if } x \textit{ is absolutely continuous} \\ \infty & \textit{else.} \end{cases}$$

Before turning to the proof of the large deviation result, we introduce the notation needed to define the rate function. Define $\mathcal{V}$ be the set of real $(I+2) \times (J+2)$ matrices, indexed over the set $\{0, \ldots, I+\} \times \{0, \ldots, J+\}$, such that the sum of all elements of each matrix is zero. Let $T^k : \mathcal{U} \to \mathcal{V}$ be defined by the expressions

$$T^1_{i,j}[\alpha] = \alpha_{i-1,j} - \alpha_{i,j} 1_{\{i \leq I\}}, \quad T^2_{i,j}[\alpha] = \alpha_{i,j-1} - \alpha_{i,j} 1_{\{j \leq J\}},$$

where for convenience we define $\alpha_{-1,j} = \alpha_{i,-1} = 0$.

Next, let $\Gamma \in \mathcal{S}$ be given with $\Gamma_{0,0}(0) = 1$. Suppose there are Borel measurable functions $\theta^k : [0,\tau] \to \mathcal{U}$, $k = 1,2$, and $x \in \mathcal{T}$ such that for all $t \in [0,\tau]$,

$$\Gamma(t) = \Gamma(0) + \int_0^t \left(\dot{x}_1 T^1[\theta^1] + \dot{x}_2 T^2[\theta^2]\right) ds. \tag{2.1}$$

Then $I(\Gamma)$ is defined by

$$I(\Gamma) = \inf_{x,\theta} \int_0^\tau \left[\dot{x}_1 R\left(\theta^1 \| \Gamma\right) + \dot{x}_2 R\left(\theta^2 \| \Gamma\right)\right] ds + J(x), \tag{2.2}$$

where the infimum is over all such θ^k and x that satisfy (2.1). If rates satisfying (2.1) exist with $I(\Gamma) < \infty$ then we say that Γ is a *valid occupancy process*. If such rates do not exist, we set $I(\Gamma) = \infty$. In Section 3 we show that for every valid occupancy path, there exist rates x and θ^k which achieve the infimum.

We interpret $\theta^k_{i,j}(t)$ as the rate at which balls of color k are thrown into cells that at time t contain i balls of color 1 and j balls of color 2, where the rates are normalized to give a probability measure for each k. We follow our usual convention that $i = I+$ refers to more than I balls, and likewise for $j = J+$. These normalized rates are modulated by the color selection process x so that $\dot{x}_k \theta^k$ represents the true rate at which balls of color k enter urns of various occupancy classes. Finally, the transformations $T^k[\alpha]$ represent the rate of change in Γ induced by balls of color k entering urns at the rates given by α.

In the next three sections, under different assumptions for the upper and lower bounds we will prove the Laplace principle for this urn model. In particular, in Section 3 we will prove

$$\liminf_{n\to\infty} -\frac{1}{n}\log E\exp\left[-nF(\Gamma^n)\right] \geq \inf_{\Gamma\in\mathcal{S}}\left[I(\Gamma)+F(\Gamma)\right],$$

and in Section 5 we will prove

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\exp\left[-nF(\Gamma^n)\right] \leq \inf_{\Gamma\in\mathcal{S}}\left[I(\Gamma)+F(\Gamma)\right].$$

These bounds are equivalent to the large deviation upper and lower bounds, respectively [3, Corollary 1.2.5]. In Section 3 we prove various properties of the rate function I, and in particular show that I has compact level sets. Although all statements and proofs are for the case of 2 colors, there are obvious extensions to the case of any finite number of colors.

To prove these bounds it will be convenient to use a representation for exponential integrals. Let $\mathcal{X}^n$ and $\mathcal{Y}^n$ denote the product space of $N_n \doteq \lfloor n\tau \rfloor + 1$ copies of $\{1,\ldots,n\}$ and $\{1,2\}$, respectively. Let Π^n denote product measure on $\mathcal{X}^n$, where each marginal of Π^n is π^n, the uniform distribution on $\{1,\ldots,n\}$, and let Λ^n denote the distribution that is induced on $\mathcal{Y}^n$ by $\{Y^n_l, l=1,\ldots,N_n\}$. Let μ^n denote any probability measure on $\mathcal{X}^n\times\mathcal{Y}^n$. Suppose that $\left\{\bar{X}^n_l, l=1,\ldots,N_n\right\}$ and $\left\{\bar{Y}^n_l, l=1,\ldots,N_n\right\}$ (on the canonical probability space $\mathcal{X}^n\times\mathcal{Y}^n$ and with expectation operator $\bar{E}^n$) have the joint distribution μ^n, and that $\bar{\Gamma}^n$ and $\bar{x}^n$ are constructed from $\left\{\bar{X}^n_l, l=1,\ldots,N_n\right\}$ and $\left\{\bar{Y}^n_l, l=1,\ldots,N_n\right\}$ in exactly the

same way that Γ^n and x^n are constructed from $\{X_l^n, l = 1, \ldots, N_n\}$ and $\{Y_l^n, l = 1, \ldots, N_n\}$. Then [3, Proposition 1.4.2]

$$-\frac{1}{n} \log E \exp\left[-nF(\Gamma^n)\right] = \inf_{\mu^n} \bar{E}^n \left[\frac{1}{n} R\left(\mu^n \| \Pi^n \otimes \Lambda^n\right) + F(\bar{\Gamma}^n)\right]. \tag{2.3}$$

The process $\bar{\Gamma}^n$ is an urn model with a "biased" or "twisted" distribution. The representation equates the normalized log of the exponential integral with a variational problem, in which we minimize the expected value of the functional F under the twisted distribution, plus a relative entropy "cost" to achieve the particular twist.

We next present an alternative expression for the relative entropy which reflects the natural relations between the underlying measures. Suppose that μ^n is decomposed into the following product of conditional distributions:

$$\begin{aligned} &\mu^n(dx_1, \ldots, dx_{N_n}, dy_1, \ldots, dy_{N_n}) \\ &\quad = \lambda^n(dy_1, \ldots, dy_{N_n}) \\ &\qquad \times \mu_{x,1}^n(dx_1 | y_1, \ldots, y_{N_n}) \cdots \mu_{x,N_n}^n\left(dx_{N_n} | x_1, \ldots, x_{N_n - 1}, y_1, \ldots, y_{N_n}\right). \end{aligned}$$

Define the random measures

$$\bar{\mu}_l^n(dx_l) \doteq \mu_{x,l}^n(dx_l | \bar{X}_r, r = 1, \ldots, l-1, \bar{Y}_r, r = 1, \ldots, N_n).$$

Thus $\bar{\mu}_l^n$ is the distribution of the cell into which the lth ball is thrown, given the outcome of all previous throws and the colors of all the balls. Using the fact that Π^n is product measure and the chain rule for relative entropy [3, Theorem C.3.1], we have

$$R\left(\mu^n \| \Pi^n \otimes \Lambda^n\right) = \bar{E}^n \left[\sum_{l=1}^{N_n} R\left(\bar{\mu}_l^n \| \pi^n\right) + R\left(\lambda^n \| \Lambda^n\right)\right]. \tag{2.4}$$

This representation separates the total relative entropy into a contribution due to the twisting of the coloration distribution, and a sum of contributions due to twisting of the distribution of the individual throws, conditioned on the coloration process and all previous throws.

3 The Large Deviation Upper Bound

In this section we prove

$$\liminf_{n \to \infty} -\frac{1}{n} \log E \exp\left[-nF(\Gamma^n)\right] \geq \inf_{\Gamma \in \mathcal{S}} \left[I(\Gamma) + F(\Gamma)\right], \tag{3.1}$$

which corresponds to the large deviation upper bound. Since in the occupation measure problem we do not distinguish between cells that contain the same number of balls of the various colors, it makes sense to rewrite the relative entropy one last time. Given $\left\{\bar{X}_r^n, l=1,\ldots,l-1\right\}$ and $\left\{\bar{Y}_l^n, l=1,\ldots,N_n\right\}$, we know that $n\bar{\Gamma}_{i,j}^n(l/n)$ is the number of cells that contain i balls of color 1 and j balls of color 2. For (i,j), let $K_{i,j}$ denote the set of cells of the corresponding type, and let $|K_{i,j}| = n\bar{\Gamma}_{i,j}^n(l/n)$ denote the number of elements of $K_{i,j}$. Let $\nu_{i,j}^n(l/n)$ denote the total probability assigned to cells of this type by $\bar{\mu}_l^n$ (the definition being irrelevant when $|K_{i,j}|=0$) :

$$\nu_{i,j}^n(l/n) = \bar{\mu}_l^n(K_{i,j}) = \sum_{m\in K_{i,j}} \bar{\mu}_l^n(\{m\}).$$

The convexity of $x\log x$ then implies the following bound:

$$\begin{aligned}
&\bar{E}^n\left[R\left(\bar{\mu}_l^n \| \pi^n\right)\right] \\
&\quad= \bar{E}^n\left[\sum_{i=0,j=0}^{I+,J+}\sum_{m\in K_{i,j}}\left[\log\left(\frac{\bar{\mu}_l^n(\{m\})}{\pi^n(\{m\})}\right)\bar{\mu}_l^n(\{m\})\right]\right] \\
&\quad\geq \bar{E}^n\left[\sum_{i=0,j=0}^{I+,J+}|K_{i,j}|\left[\log\left(\frac{\frac{\sum_{m\in K_{i,j}}\bar{\mu}_l^n(\{m\})}{|K_{i,j}|}}{\pi^n(\{m\})}\right)\frac{\sum_{m\in K_{i,j}}\bar{\mu}_l^n(\{m\})}{|K_{i,j}|}\right]\right] \\
&\quad= \bar{E}^n\left[\sum_{i=0,j=0}^{I+,J+}\left[\log\left(\frac{\nu_{i,j}^n(l/n)}{\bar{\Gamma}_{i,j}^n(l/n)}\right)\;\nu_{i,j}^n(l/n)\right]\right] \\
&\quad= \bar{E}^n\left[R\left(\nu^n(l/n)\|\bar{\Gamma}^n(l/n)\right)\right].
\end{aligned} \tag{3.2}$$

The inequality above becomes an equality when the measure $\bar{\mu}_l^n$ puts the same weight on urns of the same type, and thus one would expect this property to hold for the measure that achieves the minimum in the variational representation. For each $t\in[0,N_n/n]$ define $\nu^n(t)=\nu^n(l/n)$ if $t\in[l/n, l/n+1/n)$. Let $\hat{\Gamma}^n$ denote the piecewise constant (rather than piecewise linear) interpolant:

$$\hat{\Gamma}^n(t) = \bar{\Gamma}^n(l/n) \text{ for } t\in[l/n, l/n+1/n).$$

Note that if $\bar{\Gamma}^n$ converges uniformly to $\bar{\Gamma}$, then so does $\hat{\Gamma}^n$.

For a $\mathcal{U}$-valued process η and $x\in\mathcal{T}$, we define increasing processes $(\eta\otimes x_k)_{i,j}, i\in\{0,\ldots,I,I+\}, j\in\{0,\ldots,J,J+\}$ by

$$(\eta\otimes x_k)_{i,j}(t) = \int_0^t \eta_{i,j}(s)\dot{x}_k(s)ds.$$

When $x_k(\tau) > 0$ and $(\eta \otimes x_k)/x_k(\tau)$ appears in the relative entropy function, it is interpreted as the probability measure on $\{0, \ldots, I+\} \times \{0, \ldots, J+\} \times [0, \tau]$ that assigns to the set $A \times B$ mass

$$\int_B \sum_{i,j \in A} \eta_{i,j}(s) \dot{x}_k(s) ds \Big/ x_k(\tau).$$

Theorem 3.1 *Define the processes $\bar{\Gamma}^n, \hat{\Gamma}^n, \bar{x}_k^n, k = 1, 2$ and ν^n as above for the given measure μ^n. Then the collection*

$$\left\{ (\bar{\Gamma}^n, \bar{x}_k^n, \hat{\Gamma}^n \otimes \bar{x}_k^n, \nu^n \otimes \bar{x}_k^n), k = 1, 2, n \in \mathbb{N} \right\}$$

is tight. Thus given any subsequence there exists a further subsequence which converges in distribution to processes $\bar{\Gamma}, \bar{x}_k, \Lambda^k, \zeta^k, k = 1, 2$ defined on a probability space with expectation operator $\bar{E}$. These limit processes have the following properties.

1. *Each process $\bar{x}_k$ is absolutely continuous (w.p.1), with derivative in t denoted by $\dot{\bar{x}}_k$.*
2. *Each process ζ^k can be decomposed in the form*
$$\zeta^k = \bar{\theta}^k \otimes \bar{x}_k,$$
where the measurable process $\bar{\theta}^k$ takes values in $\mathcal{U}$.
3. *Each process Λ^k can be decomposed in the form*
$$\Lambda^k = \bar{\Gamma} \otimes \bar{x}_k.$$
4. *The relation (2.1) holds, with $\Gamma, x_1, x_2, \theta^1, \theta^2$ replaced by $\bar{\Gamma}, \bar{x}_1, \bar{x}_2, \bar{\theta}^1, \bar{\theta}^2$.*

Proof: It is easy to see that the processes $\bar{\Gamma}^n, \bar{x}_k^n, \hat{\Gamma}^n \otimes \bar{x}_k^n, \nu^n \otimes \bar{x}_k^n, k = 1, 2$ are all uniformly (in n and ω) Lipschitz continuous. Therefore the ensemble takes values in a compact set, which automatically gives tightness, and hence convergence along subsequences. If a convergent subsequence is fixed (with limit $\bar{\Gamma}, \bar{x}_k, \Lambda^k, \zeta^k, k = 1, 2$), the limit processes are also Lipschitz continuous, and hence a.e. (in t) differentiable, w.p.1. It follows directly from the definitions that $\sum_{i=0,j=0}^{I+,J+} \zeta_{i,j}^k(t) = \bar{x}_k(t)$ for $t \in [0, \tau]$. Since each component of $\zeta_{i,j}^k(t)$ is nondecreasing, there is a measurable $\mathcal{U}$-valued process $\theta_{i,j}^k$ such that $\zeta_{i,j}^k(t) = \int_0^t \theta_{i,j}^k(s) \dot{\bar{x}}_k(s) ds$. The convergence of nondecreasing

processes $\bar{x}_k^n \to \bar{x}_k$ and continuity of $\bar{\Gamma}$ imply $\bar{\Gamma} \otimes \bar{x}_k^n \to \bar{\Gamma} \otimes \bar{x}_k$. Since $\bar{\Gamma}^n$ and hence also $\hat{\Gamma}^n$ converge uniformly to $\bar{\Gamma}$, $\hat{\Gamma}^n \otimes \bar{x}_k^n \to \bar{\Gamma} \otimes \bar{x}_k$. Thus Λ_k has the indicated decomposition.

Finally we consider the last item in the theorem. Consider a component $\bar{\Gamma}_{i,j}$. We assume that $i \in \{1, \ldots, I\}, j \in \{1, \ldots, J\}$, and observe that a similar argument to the one used below will give the analogous conclusion for all other cases. Let $\mathcal{F}_l^n \doteq \sigma\left(\bar{Y}_r^n, 1 \le r \le N_n, \bar{X}_r^n, 1 \le r \le l\right)$. We can write

$$\begin{aligned}
&\bar{\Gamma}_{i,j}^n(l/n + 1/n) - \bar{\Gamma}_{i,j}^n(l/n) \\
&\quad= \frac{1}{n} 1_{\{\bar{Y}_k^n = 1\}} \Big(1_{\{\bar{X}_k^n \text{ is an urn of type } (i-1,j) \text{ at time } l\}} \\
&\qquad\qquad - 1_{\{\bar{X}_k^n \text{ is an urn of type } (i,j) \text{ at time } l\}} \Big) \\
&\qquad + \frac{1}{n} 1_{\{\bar{Y}_k^n = 2\}} \Big(1_{\{\bar{X}_k^n \text{ is an urn of type } (i,j-1) \text{ at time } l\}} \\
&\qquad\qquad - 1_{\{\bar{X}_k^n \text{ is an urn of type } (i,j) \text{ at time } l\}} \Big) \\
&\quad= \frac{1}{n} 1_{\{\bar{Y}_k^n = 1\}} \left[\nu_{i-1,j}^n(l/n) - \nu_{i,j}^n(l/n) \right] \\
&\qquad\qquad + \frac{1}{n} 1_{\{\bar{Y}_k^n = 2\}} \left[\nu_{i,j-1}^n(l/n) - \nu_{i,j}^n(l/n) \right] + e_{i,j}^n(l/n),
\end{aligned}$$

where $\left\{ e_{i,j}^n(l/n), l = 0, \ldots, N_n - 1 \right\}$ is a martingale difference with respect to $\mathcal{F}_l^n$ with $E\left[e_{i,j}^n(l/n) \right]^2 = O(1/n^2)$. Thus

$$\begin{aligned}
\bar{\Gamma}_{i,j}^n(t) - \bar{\Gamma}_{i,j}^n(0) &= (\nu^n \otimes \bar{x}_1^n)_{i-1,j}^{1,n}(t) - (\nu^n \otimes \bar{x}_1^n)_{i,j}^{1,n}(t) \\
&\quad + (\nu^n \otimes \bar{x}_2^n)_{i,j-1}^{2,n}(t) - (\nu^n \otimes \bar{x}_2^n)_{i,j}^{2,n}(t) + g_{i,j}^n(t),
\end{aligned}$$

where the process $g_{i,j}^n$ tends uniformly to zero on $[0, \tau]$. Therefore

$$\begin{aligned}
\bar{\Gamma}_{i,j}(t) - \bar{\Gamma}_{i,j}(0) &= (\theta^1 \otimes \bar{x}_1)_{i-1,j}^1(t) - (\theta^1 \otimes \bar{x}_1)_{i,j}^1(t) \\
&\quad + (\theta^2 \otimes \bar{x}_2)_{i,j-1}^2(t) - (\theta^2 \otimes \bar{x}_2)_{i,j}^2(t).
\end{aligned}$$

The last display is equivalent to the i, j-th element of (2.1). □

Theorem 3.2 *Under Assumption 2.1,*

$$\liminf_{n \to \infty} -\frac{1}{n} \log E \exp\left[-nF(\Gamma^n)\right] \ge \inf_{\Gamma \in \mathcal{S}} \left[I(\Gamma) + F(\Gamma) \right].$$

Proof: Owing to the representation, it suffices to show that

$$\liminf_{n\to\infty} \inf_{\mu^n} \bar{E}^n \left[\frac{1}{n} R\left(\mu^n \| \Pi^n \otimes \Lambda^n\right) + F(\bar{\Gamma}^n) \right] \geq \inf_{\Gamma \in \mathcal{S}} \left[I(\Gamma) + F(\Gamma) \right].$$

According to equations (2.4) and (3.2), we have the bound

$$\frac{1}{n} R\left(\mu^n \| \Pi^n \otimes \Lambda^n\right) \geq \bar{E}^n \left[\frac{1}{n} \sum_{l=1}^{N_n} R\left(\nu^n(l/n) \| \bar{\Gamma}^n(l/n)\right) + \frac{1}{n} R\left(\lambda^n \| \Lambda^n\right) \right].$$

Using the chain rule ([3, Theorem C.3.1]) again, the non-negativity of relative entropy, and $\tau \leq N_n/n$, we can write

$$\begin{aligned} &\bar{E}^n \left[\frac{1}{n} \sum_{l=1}^{N_n} R\left(\nu^n(l/n) \| \bar{\Gamma}^n(l/n)\right) \right] \\ &\quad \geq \bar{E}^n \left[\int_0^\tau R\left(\nu^n \| \hat{\Gamma}^n\right) d\bar{x}_1^n + \int_0^\tau R\left(\nu^n \| \hat{\Gamma}^n\right) d\bar{x}_2^n \right] \\ &\quad = \bar{E}^n \left[\bar{x}_1^n(\tau) R\left(\frac{\nu^n \otimes \bar{x}_1^n}{\bar{x}_1^n(\tau)} \middle\| \frac{\hat{\Gamma}^n \otimes \bar{x}_1^n}{\bar{x}_1^n(\tau)} \right) + \bar{x}_2^n(\tau) R\left(\frac{\nu^n \otimes \bar{x}_2^n}{\bar{x}_2^n(\tau)} \middle\| \frac{\hat{\Gamma}^n \otimes \bar{x}_2^n}{\bar{x}_2^n(\tau)} \right) \right]. \end{aligned}$$

According to Theorem 3.1, given any subsequence of $\mathbb{N}$ we can find a further subsequence (again denoted by n) along which we have the convergence in distribution of $(\bar{\Gamma}^n, \bar{x}_k^n, \hat{\Gamma}^n \otimes \bar{x}_k^n, \nu^n \otimes \bar{x}_k^n, k = 1, 2)$. Using Fatou's Lemma (for convergence in distribution) and the lower semicontinuity of relative entropy [3, Lemma 1.4.3],

$$\begin{aligned} &\liminf_{n\to\infty} \bar{E}^n \left[\frac{1}{n} \sum_{l=1}^{N_n} R\left(\nu^n(l/n) \| \bar{\Gamma}^n(l/n)\right) \right] \\ &\quad \geq \liminf_{n\to\infty} \bar{E}^n \left[\bar{x}_1^n(\tau) R\left(\frac{\nu^n \otimes \bar{x}_1^n}{\bar{x}_1^n(\tau)} \middle\| \frac{\hat{\Gamma}^n \otimes \bar{x}_1^n}{\bar{x}_1^n(\tau)} \right) + \bar{x}_2^n(\tau) R\left(\frac{\nu^n \otimes \bar{x}_2^n}{\bar{x}_2^n(\tau)} \middle\| \frac{\hat{\Gamma}^n \otimes \bar{x}_2^n}{\bar{x}_2^n(\tau)} \right) \right] \\ &\quad \geq \bar{E} \left[\bar{x}_1(\tau) R\left(\frac{\bar{\theta}^1 \otimes \bar{x}_1}{\bar{x}_1(\tau)} \middle\| \frac{\bar{\Gamma} \otimes \bar{x}_1}{\bar{x}_1(\tau)} \right) + \bar{x}_2(\tau) R\left(\frac{\bar{\theta}^2 \otimes \bar{x}_2}{\bar{x}_2(\tau)} \middle\| \frac{\bar{\Gamma} \otimes \bar{x}_2}{\bar{x}_2(\tau)} \right) \right] \\ &\quad = \bar{E} \left[\int_0^\tau R\left(\bar{\theta}^1(t) \| \bar{\Gamma}(t)\right) d\bar{x}_1(t) + \int_0^\tau R\left(\bar{\theta}^2(t) \| \bar{\Gamma}(t)\right) d\bar{x}_2(t) \right]. \end{aligned}$$

We claim that since $\{x^n, n \in \mathbb{N}\}$ satisfies a large deviation principle on $\mathcal{T}$ with rate function J,

$$\liminf_{n\to\infty} \frac{1}{n} R\left(\lambda^n \| \Lambda^n\right) \geq \bar{E} J(\bar{x}). \tag{3.3}$$

Indeed, it follows from the variational representation that for all bounded and continuous functions $G : \mathcal{T} \to \mathbb{R}$,

$$\liminf_{n\to\infty} \bar{E}^n \left[\frac{1}{n} R\left(\lambda^n \| \Lambda^n\right) + G(\bar{x}^n) \right] \geq \inf_{x\in\mathcal{T}} \left[J(x) + G(x) \right].$$

Thus

$$\liminf_{n\to\infty} \bar{E}^n \left[\frac{1}{n} R\left(\lambda^n \| \Lambda^n\right) \right] \geq \inf_{x\in\mathcal{T}} \left[J(x) + G(x) \right] - \bar{E}\left[G(\bar{x}) \right],$$

and since G is arbitrary,

$$\liminf_{n\to\infty} \frac{1}{n} R\left(\lambda^n \| \Lambda^n\right) \geq \sup_{G\in\mathcal{C}_b(\mathcal{T})} \left\{ \inf_{x\in\mathcal{T}} \left[J(x) + G(x) \right] - \bar{E}\left[G(\bar{x}) \right] \right\}.$$

We claim that the right hand side of this display is bounded below by $\bar{E}J(\bar{x})$. Let $-G_r$ be a sequence of bounded, non-negative continuous functions that converge up to J as $r \to \infty$. It follows that $J(x) + G_r(x) \geq 0$ for all r and $x \in \mathcal{T}$, and so $\inf_{x\in\mathcal{T}}[J(x) + G_r(x)] \geq 0$. Since the monotone convergence theorem implies $\bar{E}\left[-G_r(\bar{x})\right] \uparrow \bar{E}\left[J(\bar{x})\right]$, the result now follows.

We have the following inequalities, each of which is explained after the display.

$$\begin{aligned}
&\liminf_{n\to\infty} \; -\frac{1}{n} \log E \exp\left[-nF(\Gamma^n)\right] \\
&\quad = \liminf_{n\to\infty} \bar{E}^n \left[\frac{1}{n} R\left(\mu^n \| \Pi^n \otimes \Lambda^n\right) + F(\bar{\Gamma}^n) \right] \\
&\quad \geq \liminf_{n\to\infty} \bar{E}^n \left[\frac{1}{n} \sum_{l=1}^{N_n} R\left(\nu^n(l/n) \| \bar{\Gamma}^n(l/n)\right) + \frac{1}{n} R\left(\lambda^n \| \Lambda^n\right) + F(\bar{\Gamma}^n) \right] \\
&\quad \geq \bar{E}\left[\int_0^{\mathcal{T}} R\left(\bar{\theta}^1(t) \left\| \bar{\Gamma}(t)\right.\right) d\bar{x}_1(t) \right. \\
&\qquad\qquad \left. + \int_0^{\mathcal{T}} R\left(\bar{\theta}^2(t) \left\| \bar{\Gamma}(t)\right.\right) d\bar{x}_2(t) + J(\bar{x}) + F(\bar{\Gamma}) \right] \\
&\quad \geq \bar{E}\left[I(\bar{\Gamma}) + F(\bar{\Gamma}) \right] \\
&\quad \geq \inf_{\Gamma\in\mathcal{S}} \left[I(\Gamma) + F(\Gamma) \right].
\end{aligned}$$

The first equality is due to the relative entropy representation and the fact that μ^n is a minimizer; the following inequality uses the decomposition (2.4) and the bound (3.2); the second inequality uses the bound (3.3), the bound

immediately above (3.3), the convergence in distribution of $\bar{\Gamma}^n$ to $\bar{\Gamma}$, and the continuity of F; the third inequality uses the properties of the limit processes stated in Theorem 3.1 and the definition of the rate function; the final inequality is obvious. We have proved that given any subsequence of $\mathbb{N}$ there is a further subsequence along which (3.1) holds. By the usual argument by contradiction, (3.1) holds as stated. $\square$

4 Properties of the Rate Function

In this section we prove some important properties of the rate function.

Theorem 4.1 *Under Assumption 2.1 the set $\{\Gamma : I(\Gamma) \leq M\}$ is compact for each $M \in [0, \infty)$.*

Proof: Since all paths Γ with $I(\Gamma) < \infty$ are Lipschitz continuous with a common constant, we need only show that $\Gamma \to I(\Gamma)$ is lower semicontinuous. Let $\Gamma_n \to \Gamma$ as $n \to \infty$. If $\liminf_{n\to\infty} I(\Gamma_n) < I(\Gamma)$, then we can extract a subsequence (again denoted by n) such that $I(\Gamma_n)$ converges and $\lim_{n\to\infty} I(\Gamma_n) = I(\Gamma) - \varepsilon$ for some $\varepsilon > 0$. Let $\theta^{k,n}, k = 1, 2$ and x^n be associated rates and cumulative coloration processes that satisfy

$$\dot{\Gamma}_n = \dot{x}_1^n T^1 \left[\theta^{1,n}\right] + \dot{x}_2^n T^2 \left[\theta^{2,n}\right]$$

and

$$I(\Gamma_n) = \int_0^\tau \left[\dot{x}_1^n R(\theta^{1,n} \,\|\Gamma) + \dot{x}_2^n R(\theta^{2,n} \,\|\Gamma)\right] dt + J(x^n) + 1/n.$$

Exactly as in the proof of the convergence theorem (Theorem 3.1), the uniformly Lipschitz continuous processes $(\Gamma_n, x_k^n, \theta^{k,n} \otimes x_k^n, \Gamma_n \otimes x_k^n, k = 1, 2)$ converge, at least along a subsequence, to a collection of processes $(\Gamma, x_k, \theta^k \otimes x_k, \Gamma \otimes x_k, k = 1, 2)$. Using the lower semicontinuity of J and the relative entropy,

$$\begin{aligned}
&\liminf_{n\to\infty} \left(\int_0^\tau \left[\dot{x}_1^n R(\theta^{1,n} \,\|\Gamma_n) + \dot{x}_2^n R(\theta^{2,n} \,\|\Gamma_n)\right] dt + J(x^n) \right) \\
&= \liminf_{n\to\infty} \left(\left[x_1^n(\tau) R \left(\frac{\theta^{1,n} \otimes x_1^n}{x_1^n(\tau)} \,\middle\|\, \frac{\Gamma_n \otimes x_1^n}{x_1^n(\tau)} \right) \right.\right. \\
&\qquad \left.\left. + x_2^n(\tau) R \left(\frac{\theta^{2,n} \otimes x_2^n}{x_2^n(\tau)} \,\middle\|\, \frac{\Gamma_n \otimes x_2^n}{x_2^n(\tau)} \right) \right] dt + J(x^n) \right)
\end{aligned}$$

$$
\begin{aligned}
\geq \quad & \left(\left[x_1(\tau) R \left(\frac{\theta^1 \otimes x_1}{x_1(\tau)} \,\middle\|\, \frac{\Gamma \otimes x_1}{x_1(\tau)} \right) \right.\right. \\
& \left.\left. + x_2(\tau) R \left(\frac{\theta^2 \otimes x_2}{x_2(\tau)} \,\middle\|\, \frac{\Gamma \otimes x_2}{x_2(\tau)} \right) \right] dt + J(x) \right) \\
= \quad & \left(\int_0^\tau \left[\dot{x}_1 R(\theta^1 \,\|\Gamma) + \dot{x}_2 R(\theta^2 \,\|\Gamma) \right] dt + J(x) \right).
\end{aligned}
$$

Since we also have

$$
\begin{aligned}
\Gamma(t) - \Gamma(0) &= \lim_{n\to\infty} \left(\Gamma_n(t) - \Gamma_n(0) \right) \\
&= \lim_{n\to\infty} \int_0^t \left[\dot{x}_1^n T^1 \left[\theta^{1,n} \right] + \dot{x}_2^n T^2 \left[\theta^{2,n} \right] \right] ds \\
&= \lim_{n\to\infty} \left(T^1 \left[\int_0^t \dot{x}_1^n \theta^{1,n} ds \right] + T^2 \left[\int_0^t \dot{x}_2^n \theta^{2,n} ds \right] \right) \\
&= T^1 \left[\int_0^t \dot{x}_1 \theta^1 ds \right] + T^2 \left[\int_0^t \dot{x}_2 \theta^2 ds \right] \\
&= \int_0^t \left[\dot{x}_1 T^1 \left[\theta^1 \right] + \dot{x}_2 T^2 \left[\theta^2 \right] \right] ds,
\end{aligned}
$$

we conclude that

$$
I(\Gamma) \leq \left(\int_0^\tau \left[\dot{x}_1 R(\theta^1 \,\|\Gamma) + \dot{x}_2 R(\theta^2 \,\|\Gamma) \right] dt + J(x) \right) \leq \lim_{n\to\infty} I\left(\Gamma_n\right),
$$

a contradiction. Therefore, $\liminf_{n\to\infty} I\left(\Gamma_n\right) \geq I(\Gamma)$. □

It is also true that given any Γ, infimizing θ's and x's exist.

Lemma 4.2 *Let $\Gamma \in \mathcal{S}$ be given. There exist measurable functions $\theta^k, k = 1, 2$ and $x \in \mathcal{T}$ which achieve the infimum in the definition of $I(\Gamma)$.*

Proof: Since the proof uses the same ideas as that of the previous theorem, the argument is only sketched. If $I(\Gamma) = \infty$ there is nothing to prove. If $I(\Gamma) < \infty$ then there exist $\theta^{k,n}, k = 1, 2$ and $x^n \in \mathcal{T}$ such that

$$
\dot{\Gamma} = \dot{x}_1^n T^1 \left[\theta^{1,n} \right] + \dot{x}_2^n T^2 \left[\theta^{2,n} \right]
$$

and

$$
\int_0^\tau \left[\dot{x}_1^n R\left(\theta^{1,n} \,\|\Gamma \right) + \dot{x}_2^n R\left(\theta^{2,n} \,\|\Gamma \right) \right] dt \leq I(\Gamma) + 1/n.
$$

Arguing exactly as in the previous theorem, we can consider the limit $n \to \infty$ (along a subsequence) and construct the minimizing $\theta^k, k = 1, 2$ and $x \in \mathcal{T}$. □

A special but common case is when the coloration rate function takes the form

$$J(x) = \int_0^\tau M(\dot{x}_1, \dot{x}_2) dt$$

for some convex proper function $M : \mathbb{R}^2 \to [0, \infty]$ (see the examples of Section 1). In this case we can write I as

$$I(\Gamma) = \int_0^\tau L(\Gamma, \dot{\Gamma}) dt.$$

in terms of a local rate function

$$\begin{aligned} L(\Gamma, \eta) \doteq \inf \big\{ & a_1 R(\theta^1 \, \|\Gamma) + a_2 R(\theta^2 \, \|\Gamma) + M(a_1, a_2) : \\ & a \in \mathcal{C},\ \theta^k \in \mathcal{U},\ \eta = a_1 T^1 \left[\theta^1\right] + a_2 T^2 \left[\theta^2\right] \big\}, \end{aligned} \quad (4.1)$$

where $\mathcal{C}$ is the set of probability distributions on $\{1, 2\}$. We will rely on the following assumption in the proof of the lower bound.

Assumption 4.1 *Assumption 2.1 holds, and in addition:*

(A) The rate function $J(x)$ takes the form $\int_0^\tau M(\dot{x}) dt$, where M is a proper convex function.

(B) There is a point $a \in \mathcal{C}$ such that $M(a) = 0$ and $a_i > 0, i = 1, 2$.

As noted previously, the assumed form for J is typical. Since M is a rate function, there is at least one probability vector a at which $M(a) = 0$. The assumption that this occurs at a point where both components are positive is very mild.

In the remainder of this section we will construct processes and controls that will be used in the proof of the large deviation lower bound. We first define the natural occupancy path corresponding to a given colorization process, and the zero-cost path. For $y \in [0, \infty)$ and $i \in \mathbb{Z}$ let

$$\mathcal{P}_i(y) = \begin{cases} \frac{y^i}{i!} e^{-y} & i \geq 0 \\ 0 & i < 0 \end{cases}$$

denote the i-th component of a Poisson distribution with mean y, and let $\mathcal{Q}_i(y) = \sum_{j>i} \mathcal{P}_j(y)$ be the Poisson tail probability function. The product of two independent Poisson distributions with means y_1 and y_2 defines a mapping $\Phi : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathcal{U}$ given by

$$\begin{aligned}
\Phi_{i,j}(y_1, y_2) &= \mathcal{P}_i(y_1)\mathcal{P}_j(y_2) \\
\Phi_{i,J+}(y_1, y_2) &= \mathcal{P}_i(y_1)\mathcal{Q}_J(y_2) \\
\Phi_{I+,j}(y_1, y_2) &= \mathcal{Q}_I(y_1)\mathcal{P}_j(y_2) \\
\Phi_{I+,J+}(y_1, y_2) &= \mathcal{Q}_I(y_1)\mathcal{Q}_J(y_2).
\end{aligned}$$

The mapping Φ is the limiting (as $n \to \infty$) mean urn occupancy distribution for an experiment in which ny_k balls of color k are thrown into n urns.

Lemma 4.3 (Natural occupancy path) *Suppose that condition (A) of Assumption 4.1 holds. Let $x \in \mathcal{T}$ be a colorization process with finite cost $J(x)$. The* natural occupancy path *corresponding to x defined by $\Gamma^*(t) = \Phi(x_1(t), x_2(t))$ is an occupancy path in $\mathcal{S}$ which satisfies the initial condition $\Gamma^*_{0,0}(0) = 1$ and the bound $I(\Gamma^*) \leq J(x)$.*

Proof: The initial condition is immediate from the fact that $x(0) = 0$. The continuity of x and of the Poisson distribution with respect to its mean ensure that $\Gamma^* \in \mathcal{S}$. To establish the bound on $I(\Gamma^*)$, we will show that the derivative of the path satisfies the differential equation

$$\dot{\Gamma}^* = \dot{x}_1 T^1[\Gamma^*] + \dot{x}_2 T^2[\Gamma^*].$$

Then according to the definition of the rate function,

$$I(\Gamma^*) \leq \int_0^\tau \left(\sum_{k=1}^{2} \dot{x}_k R\left(\Gamma^* \mid\mid \Gamma^*\right) + M(\dot{x}) \right) dt = \int_0^\tau M(\dot{x})\, dt = J(x).$$

Note that $\frac{d}{dx}\mathcal{P}_i(x) = \mathcal{P}_{i-1}(x) - \mathcal{P}_i(x)$. Then, for $i \leq I$ and $j \leq J$, we have

$$\begin{aligned}
\dot{\Gamma}^*_{i,j} &= \dot{x}_1 \left[\mathcal{P}_{i-1}(x_1) - \mathcal{P}_i(x_1)\right] \mathcal{P}_j(x_2) + \dot{x}_2 \mathcal{P}_i(x_1) \left[\mathcal{P}_{j-1}(x_2) - \mathcal{P}_j(x_2)\right] \\
&= \dot{x}_1 \left(\Gamma^*_{i-1,j} - \Gamma^*_{i,j}\right) + \dot{x}_2 \left(\Gamma^*_{i,j-1} - \Gamma^*_{i,j}\right)
\end{aligned}$$

as desired. Using the fact that $\frac{d}{dx}\mathcal{Q}_i(x) = \mathcal{P}_i(x)$, the cases involving $i = I+$ or $j = J+$ follow similarly. $\square$

The following lemma is immediate, using the linear colorization process $x(t) = at$.

Lemma 4.4 (Zero cost path) *Suppose that Assumption 4.1 holds, with $a \in \mathcal{C}$ such that $M(a) = 0$. Then the function $Z(t) = \Phi(a_1 t, a_2 t)$ is an occupancy path in $\mathcal{S}$ which satisfies $I(Z) = 0$ and the initial condition $Z_{0,0}(0) = 1$.*

In the single color case analyzed in [4], the local rate function expressed by (4.1) takes the simpler form

$$L_s(\Gamma, \eta) = R(\Theta \,||\, \Gamma)$$

where Γ and Θ are probability distributions on $\{0, \ldots, I+\}$ and where Θ has the explicit form $\Theta_i = -\sum_{j=0}^{i} \eta_j$. In that case, the convexity of the relative entropy implies that the local rate function is a convex function of its arguments.

In the present case, the local rate function is a convex function of η but is not necessarily jointly convex in (Γ, η). It can be shown in fact that, under Assumption 4.1, L is convex if and only if the function $M(a)+h(a)$ is convex for $a \in \mathcal{C}$, where $h(a) = -a_1 \log a_1 - a_2 \log a_2$ is the entropy function. It is easy to see that Examples 2.1 and 2.2 always satisfy this condition. However, it can also be demonstrated that Example 2.3 satisfies this condition if and only if the Markov transition probabilities satisfy $p_{11} + p_{22} \leq 1$.

In our proof of the lower bound, we require a technical result showing that every valid occupancy path Γ is close, both in sup norm and in cost, to an occupancy path for which each element is bounded away from zero by a power of t. This fact allows us to avoid explictly considering occupancy paths which are close to the boundary after time $t = 0$. When the local rate function is convex, this technical result is easily demonstrated by slightly perturbing the given occupancy path in the direction of the zero-cost path (which is itself avoids the boundary after $t = 0$).

We will use a modified form of this argument to establish this result without requiring convexity of the local rate function. We first show that Γ is close to an occupancy path $\hat{\Gamma}$ whose optimal colorization process $\hat{x}$ has each component $\hat{x}_k(t)$ bounded below by a function of the form δt for some $\delta > 0$. Secondly, we show that perturbing $\hat{\Gamma}$ in the direction of the natural occupancy process Γ^* corresponding to $\hat{x}$ yields an occupancy process with the desired properties. These steps are undertaken in the next two lemmas.

Lemma 4.5 *Suppose that Assumption 4.1 holds. Let $\Gamma \in \mathcal{S}$ be given and let $\varepsilon > 0$. There exists $\hat{\Gamma} \in \mathcal{S}$, with $I(\hat{\Gamma}) < \infty$ and associated optimal rates $\hat{x}$ and $\hat{\theta}$, which further satisfies the following properties:*

- $I(\hat{\Gamma}) \leq I(\Gamma) + \varepsilon$,

- *$d(\Gamma, \hat{\Gamma}) \leq \varepsilon$,*
- *there exists $\delta > 0$ such that $\hat{x}_k(t) \geq \delta t$ for all $t \in [0, \tau]$.*

Proof: We will construct $\hat{\Gamma}$ by following the zero cost path Z of Lemma 4.4 for a short time Δ and then closely tracking the original process Γ for $t > \Delta$.

The key step is to construct an occupancy process $\tilde{\Gamma}$ which has initial condition $\tilde{\Gamma}(0) = Z(\Delta)$, which for sufficiently small Δ satisfies $d(\Gamma, \tilde{\Gamma}) \leq \varepsilon/2$ and $I(\tilde{\Gamma}) \leq I(\Gamma) + \varepsilon$.

Once $\tilde{\Gamma}$ has been constructed, we then may define

$$\hat{\Gamma}(t) = \begin{cases} Z(t) & 0 \leq t \leq \Delta \\ \tilde{\Gamma}(t - \Delta) & \Delta < t \leq \tau. \end{cases}$$

Since the absolute value of the derivative of each element of an occupancy function with finite cost is bounded by 1, we immediately have $d(\hat{\Gamma}, \tilde{\Gamma}) \leq \Delta$, and hence $d(\hat{\Gamma}, \Gamma) < \varepsilon$ for sufficiently small Δ. Moreover

$$I(\hat{\Gamma}) = 0 + \int_0^{\tau - \Delta} L\left(\tilde{\Gamma}, \dot{\tilde{\Gamma}}\right) dt \leq I(\tilde{\Gamma}) \leq I(\Gamma) + \varepsilon.$$

Finally, the optimal colorization rate for $\hat{\Gamma}$ on the interval $[0, \Delta]$ is given by $\hat{x}_k(t) = a_k t$. Because the colorization processes are monotonically increasing, we have the bound $\hat{x}_k(t) \geq (a_k(\Delta \wedge \tau)/\tau)\ t$ for all $t \in [0, \tau]$

It remains to construct an occupancy function $\tilde{\Gamma}$ with the required properties. We define

$$\tilde{\Gamma}(t) = e^{-\Delta}\left(\Gamma(t) - \Gamma(0)\right) + Z(\Delta) = e^{-\Delta}\Gamma(t) + (1 - e^{-\Delta})Z_c(\Delta)$$

where $Z(\Delta)$ is the zero-cost distribution of Lemma 4.4 at time Δ, and where

$$Z_c(\Delta) = \frac{Z(\Delta) - e^{-\Delta}\Gamma(0)}{1 - e^{-\Delta}}$$

is the conditional distribution obtained from $Z(\Delta)$ by removing the probability mass from its $(0, 0)$ element. As $\tilde{\Gamma}$ is a convex combination of two elements of $\mathcal{S}$, we immediately have $\tilde{\Gamma} \in \mathcal{S}$, and it is also clear that $d(\tilde{\Gamma}, \Gamma)$ can be made arbitrarily small by decreasing Δ.

It remains to establish the desired bound on $I(\tilde{\Gamma})$. If $\alpha \in \mathcal{U}$ is the distribution that puts all of its mass on the $(I+, J+)$ element, note that $T^k[\alpha] = 0$ for $k = 1, 2$, reflecting that balls thrown into $I+, J+$ urns have

no effect on the occupancy state. Let x and θ be the optimal rate processes for Γ. Then the rates $\tilde{x} = x$ and $\tilde{\theta}^k = e^{-\Delta}\theta^k + (1-e^{-\Delta})\alpha$ satisfy

$$\dot{\tilde{x}}_1 T^1[\tilde{\theta}^1] + \dot{\tilde{x}}_2 T^2[\tilde{\theta}^2] = e^{-\Delta}\left(\dot{x}_1 T^1[\theta^1] + \dot{x}_2 T^2[\theta^2]\right) = e^{-\Delta}\dot{\Gamma} = \dot{\tilde{\Gamma}}$$

and are therefore feasible rates for constructing $\tilde{\Gamma}$. Using $\tilde{x} = x$ and the joint convexity of the relative entropy, we have

$$\begin{aligned} I(\tilde{\Gamma}) &\leq \int_0^\tau \sum_{k=1}^2 \dot{\tilde{x}}_k R\left(\tilde{\theta}^k \parallel \tilde{\Gamma}\right) dt + J(\tilde{x}) \\ &\leq e^{-\Delta}\int_0^\tau \sum_{k=1}^2 \dot{x}_k R\left(\theta^k \parallel \Gamma\right) dt \\ &\quad + (1-e^{-\Delta})\int_0^\tau \sum_{k=1}^2 \dot{x}_k R\left(\alpha \parallel Z_c(\Delta)\right) dt + J(x) \\ &\leq I(\Gamma) + \tau(1-e^{-\Delta}) R\left(\alpha \parallel Z_c(\Delta)\right). \end{aligned}$$

Using the bound $Z_{I+,J+}(\Delta) \geq \mathcal{P}_{I+1}(a_1\Delta)\mathcal{P}_{J+1}(a_2\Delta)$ the difference $I(\tilde{\Gamma}) - I(\Gamma)$ is bounded by

$$\tau(1-e^{-\Delta})\log\left(a_1^{-I-1}a_2^{-J-1}\Delta^{-(I+J+2)}(I+1)!(J+1)!e^{\Delta}(1-e^{-\Delta})\right),$$

which approaches zero as $\Delta \to 0$. □

Lemma 4.6 *Suppose that Assumption 4.1 holds. Let $\Gamma \in \mathcal{S}$ be given such that $I(\Gamma) < \infty$, and let $\varepsilon > 0$. Then there exists $\delta > 0, K \in \mathbb{N}$ and $\Gamma^\varepsilon \in \mathcal{S}$ with the following properties.*

1. *$I(\Gamma^\varepsilon) \leq I(\Gamma) + \varepsilon$,*
2. *$d(\Gamma^\varepsilon, \Gamma) \leq \varepsilon$,*
3. *$\Gamma^\varepsilon_{i,j}(t) > \delta t^K$ for all $t \in (0,\tau]$ and $(i,j) \in \{0,1,\ldots,I,I+\} \times \{0,1,\ldots,J,J+\}$.*

Proof: Using Lemma 4.5, there exists $\hat{\Gamma}$ and $\bar{\delta} > 0$ with $d(\Gamma, \hat{\Gamma}) < \varepsilon/2$, $I(\hat{\Gamma}) \leq I(\Gamma) + \varepsilon$, and $\hat{x}_k \geq \bar{\delta} t$, where $\hat{x}$ and $\hat{\theta}$ are the optimal rates used in the definition of $I(\hat{\Gamma})$. Let $\Gamma^* = \Phi(\hat{x}_1, \hat{x}_2)$ be the natural occupancy path corresponding to $\hat{x}$, and define

$$\Gamma^\varepsilon = (1-\lambda)\hat{\Gamma} + \lambda\Gamma^*.$$

By the triangle inequality, Γ^ε will be sufficiently close to Γ if $\lambda < \varepsilon/4$. Note that the rates $x^\varepsilon = \hat{x}$ and $\theta^{k,\varepsilon} = (1-\lambda)\hat{\theta}^k + \lambda\Gamma^*$ are feasible rates for generating $\dot{\Gamma}^\varepsilon$, so that

$$\begin{aligned}
I(\Gamma^\varepsilon) &\leq \int_0^\tau \sum_{k=1}^2 \dot{\hat{x}}_k R\left(\theta^{k,\varepsilon} \mid\mid \Gamma^\varepsilon\right) dt + J(\hat{x}) \\
&\leq (1-\lambda) I(\hat{\Gamma}) + \lambda I(\Gamma^*) \\
&\leq I(\hat{\Gamma}) \\
&\leq I(\Gamma) + \varepsilon
\end{aligned}$$

where we have used convexity of the relative entropy and the fact that $I(\Gamma^*) = J(\hat{x}) \leq I(\hat{\Gamma})$.

Finally, for $i \leq I, j \leq J$, the i,j-th element of Γ^ε satisfies the lower bound

$$\begin{aligned}
\Gamma^\varepsilon_{i,j}(t) &\geq \lambda \Gamma^*_{i,j}(t) \\
&= \lambda \frac{\hat{x}_1^i(t)\hat{x}_2^j(t)}{i!j!} e^{-t} \\
&\geq \lambda \frac{\bar{\delta}^{i+j} e^{-\tau}}{i!j!} t^{i+j}
\end{aligned}$$

for all $t \in [0,\tau]$. Similar bounds hold for cases involving $i = I+$ and $j = J+$, since $\mathcal{Q}_i \geq \mathcal{P}_{i+1}$. Hence Γ^ε has all of the desired properties. $\square$

The final lemma of this section is the principle result that will be used to support the proof of the lower bound.

Lemma 4.7 *Suppose that Assumption 4.1 holds. Let $\Gamma \in \mathcal{S}$ be given such that $I(\Gamma) < \infty$ and such that for some $\delta > 0$ and $K \in \mathbb{N}$ the lower bound $\Gamma_{i,j}(t) \geq \delta t^K$ holds for all $(i,j) \in \{0,1,\ldots,I,I+\} \times \{0,1,\ldots,J,J+\}$. Let x, θ^1 and θ^2 satisfy*

$$\dot{\Gamma} = \dot{x}_1 T^1\left[\theta^1\right] + \dot{x}_2 T^2\left[\theta^2\right]$$

and

$$I(\Gamma) = \int_0^\tau \left[\dot{x}_1 R\left(\theta^1 \| \Gamma\right) + \dot{x}_2 R\left(\theta^2 \| \Gamma\right)\right] dt + J(x).$$

Given $\varepsilon > 0$ there exist $\Gamma^*, \theta^{1,*}, \theta^{2,*}$ and $\sigma > 0$ with the following properties.

1.

$$\dot{\Gamma}^* = \dot{x}_1 T^1\left[\theta^{1,*}\right] + \dot{x}_2 T^2\left[\theta^{2,*}\right], \Gamma^*_{0,0}(0) = 1,$$

2.
$$I(\Gamma^*) \le \int_0^\tau \left[\dot{x}_1 R\left(\theta^{1,*} \left\| \Gamma^*_{i,j}\right.\right) + \dot{x}_2 R\left(\theta^{2,*} \left\| \Gamma^*_{i,j}\right.\right)\right] dt + J(x) \le I(\Gamma) + \varepsilon,$$

3. $d(\Gamma^*, \Gamma) \le \varepsilon$,

4. *The rate processes* $\theta^{1,*}$ *and* $\theta^{2,*}$ *are piecewise constant on* $[0,\tau]$*, with a finite number of intervals of constancy.*

5. *When restricted to* $[0,\sigma)$*, the rate processes are pure in the sense that on any interval of constancy* (s_1, s_2)*, and for each* $k = 1, 2$*, there is* $(i,j) \in \{0, 1, \ldots, I, I+\} \times \{0, 1, \ldots, J, J+\}$ *such that* $\theta^{k,*}_{i,j}(t) = 1$ *for all* $t \in (s_1, s_2)$*. In addition, for* $k = 1, 2$ *and any interval of constancy on which* $\theta^{k,*}_{i,j}(t) = 1$*,* $\Gamma^*_{i,j}(t) \ge \delta\sigma^K$*.*

Proof: Suppose that $\Gamma, \theta^1, \theta^2$, and x satisfy the assumptions of the lemma. Let $\sigma \in (0,\tau]$. By assumption, we have $\Gamma_{i,j}(\sigma) \ge \delta\sigma^K$. We can choose $\sigma > 0$ such that $-\sigma \log\left(\delta\sigma^K\right) \le \varepsilon/2$, and such that if Γ_1 and Γ_2 are any occupancy processes with the same initial condition, then $\|\Gamma_1(s) - \Gamma_2(s)\| \le \varepsilon$ for all $s \in [0,\sigma]$ (here we use the common Lipschitz continuity for all occupancy functions). A time-reversed induction argument will be used to construct the pure controls on $[0,\sigma)$ described in the lemma. The main idea is that the occupancy path Γ^* will proceed on $[0,\sigma]$ in such a way that each element increases to a maximum level before decreasing to its final value $\Gamma^*(\sigma) = \Gamma(\sigma)$. Hence the contents of any given occupancy class are *only reduced at times when that class has at least a fraction* $\delta\sigma^K$ *of the urns*. In other words, $\theta^{k,*}_{i,j}(s) > 0$ implies $\Gamma^*_{i,j}(s) \ge \delta\sigma^K$, allowing the cost of Γ^* on $(0,\sigma)$ to be made arbitrarily small.

For a given $\Gamma \in \mathcal{U}$, the associated minimum number of balls of each color per urn is given by applying the linear operators

$$\begin{aligned} \beta^e_1(\Gamma) &= \sum_{i=0}^{I}\sum_{j=0}^{J+} i\Gamma_{i,j} + \sum_{j=0}^{J+}(I+1)\Gamma_{I+,j} \\ \beta^e_2(\Gamma) &= \sum_{j=0}^{J}\sum_{i=0}^{I+} j\Gamma_{i,j} + \sum_{i=0}^{I+}(J+1)\Gamma_{i,J+}. \end{aligned} \tag{4.2}$$

Because x is a coloring process which generates $\Gamma(\sigma)$ it must satisfy $x_k(\sigma) \ge \beta^e_k(\Gamma(\sigma))$. As another way to see this point, one may verify the relations

$$\begin{array}{ll} \beta^e_1(T^1[\theta]) = \sum_{i=0}^{I}\sum_{j=0}^{J+}\theta_{i,j} & \beta^e_1(T^2[\theta]) = 0 \\ \beta^e_2(T^1[\theta]) = 0 & \beta^e_2(T^2[\theta]) = \sum_{i=0}^{I+}\sum_{j=0}^{J}\theta_{i,j}. \end{array} \tag{4.3}$$

The definition of $\dot{\Gamma}$ in terms of $\dot{x}, \theta$ then implies that $0 \leq \dot{\beta}^e_k(\Gamma) \leq \dot{x}_k$.

To simplify the exposition, we first assume $x_k(\sigma) = \beta^e_k(\Gamma(\sigma))$, and subsequently extend the argument to cover inequality.

For colors $k = 1, 2$, we define orderings $\xi_k : \{0, \ldots, I+\} \times \{0, \ldots, J+\} \to N$ by

$$\xi_1(i,j) = (J+1)i + j, \quad \xi_2(i,j) = i + (I+1)j,$$

where strictly speaking we substitute $i = I+1$ or $j = J+1$ on the right-hand side when $i = I+$ or $j = J+$ appears on the left. Observe that each class of urns corresponds to a distinct value of ξ_k. For any $\Gamma \in \mathcal{U}$, let $U(\Gamma)$ be the set of (i,j) pairs such that $\Gamma_{i,j} > 0$, and for $k = 1, 2$ define

$$\begin{aligned} \kappa_k(\Gamma) &= \max_{U(\Gamma)} \xi_k(i,j) \\ u_k(\Gamma) &= \arg\max_{U(\Gamma)} \xi_k(i,j). \end{aligned}$$

To begin the induction, we set $\tau_1 = \sigma$ and $\Gamma^*(\tau_1) = \Gamma(\sigma)$, noting that $x_k(\tau_1) = \beta^e_k(\Gamma^*(\tau_1))$, and initialize the induction variable as $m = 1$.

Denote the highest non-empty urn class under the color-k ordering by $u^m_k = u_k(\Gamma^*(\tau_m))$, and denote the corresponding order number by $\kappa^m_k = \xi_k(u^m_k)$. Now imagine in reverse time pulling balls from these urns so that mass drains from $\Gamma_{u^m_1}$ at a rate specified by $\dot{x}_1$ and from $\Gamma_{u^m_2}$ at a rate specified by $\dot{x}_2$. At some time $\tau_{m+1} < \tau_m$, one of the two urn classes will empty. If $u^m_1 = u^m_2$, this time is given by

$$\tau_{m+1} = \max\left\{ t : \sum_{k=1}^{2} x_k(\tau_m) - x_k(t) = \Gamma^*_{u^1_1}(\tau_m) \right\}$$

and otherwise it is given by

$$\tau_{m+1} = \max\{t : x_1(\tau_m) - x_1(t) = \Gamma^*_{u^m_1}(\tau_m) \text{ or } x_2(\tau_m) - x_2(t) = \Gamma^*_{u^m_2}(\tau_m)\}.$$

Suppose that $u^m_1 = (i,j)$ for some $i > 0$. Then (see (4.2)),

$$x_1(\tau_m) = \beta^e_1(\Gamma^*(\tau_m)) \geq \Gamma^*_{u^m_1}(\tau_m)$$

so that there must be a solution $t \in [0, \tau_m)$ to $x_1(t) = x_1(\tau_m) - \Gamma^*_{u^m_1}(\tau_m)$ Together with similar analysis of u^m_2, this means that a solution $\tau_{m+1} \in [0, \tau_m)$ always exists unless $u^m_1 = u^m_2 = (0,0)$. In the latter case, all urns are empty at time τ_m, and $\kappa^m_1 + \kappa^m_2 = 0$.

Otherwise, if $\kappa^m_1 + \kappa^m_2 > 0$, we define

$$\theta^{1,*}(t) = e_{\{u^m_1 - (1,0)\}} \text{ and } \theta^{2,*}(t) = e_{\{u^m_2 - (0,1)\}}$$

for all $t \in [\tau_{m+1}, \tau_m)$, where e_{ij} is the distribution in $\mathcal{U}$ which puts all mass on the i, j-th element. That is, during the m-th interval, in forward time, we are throwing balls into urns with occupancies $u_1^m - (1,0)$ and $u_2^m - (0,1)$ in order to fill the urn classes u_k^m. The process Γ^* is then defined on the interval $[\tau_{m+1}, \tau_m)$ by the existing terminal condition $\Gamma^*(\tau_m)$ and the differential equation $\dot{\Gamma}^* = \sum_k \dot{x}_k T^k[\theta^{k,*}]$. Because the $\theta^{k,*}$ put all their mass on urn classes with $i \leq I, j \leq J$, the relations (4.3) establish that $x(\tau_{m+1}) = \beta_k^e(\Gamma^*(\tau_{m+1}))$, setting up the next induction step.

Note that τ_{m+1} was chosen so that at least one of the classes u_k^m is empty in $\Gamma^*(\tau_{m+1})$. This ensures the strict inequality $\sum_{k=1}^2 \kappa_k^{m+1} < \sum_{k=1}^2 \kappa_k^m$, so that the induction must terminate after a finite number of steps (say M) with $\kappa_1^M + \kappa_2^M = 0$, meaning that $\Gamma^*(\tau_M) = e_{0,0} = \Gamma(0)$. Moreover, the fact that $x_k(\tau_M) = \beta_k^e(\Gamma^*(\tau_M)) = 0$ shows that this occurs at time $\tau_M = 0$.

This establishes the existence of Γ^*, θ^* on $[0, \sigma)$ with $\Gamma^*(\sigma) = \Gamma(\sigma)$ and with θ^* consisting of pure, piecewise constant controls. In addition, $\Gamma_{i,j}^*(t)$ increases during intervals when $u_1^m = (i-1, j)$ or $u_2^m = (i, j-1)$, and it decreases when $u_k^m = (i,j)$ for $k = 1, 2$. By construction, the order numbers κ_k^m increase monotonically with decreasing m, ensuring that the intervals on which $\Gamma_{i,j}^*$ increases preceed the intervals of decrease.

Finally, we extend the argument to the case when $x_k(\sigma) > \beta_k^e(\Gamma(\sigma))$ for some $k \in 1, 2$. Let s_k be the last time in $[0, \sigma]$ such that $x_k(s_k) = \beta_k^e(\Gamma(\sigma))$, and assume without loss that $0 < s_2 < s_1 < \sigma$. By virtue of (4.3), the event $x_1(t) > \beta_1^e(\Gamma(t))$ can only occur if balls of class 1 have been thrown into $(I+, j)$-occupied urns, which can only happen with finite cost if these urns hold some mass: $\sum_{j=0}^{J+} \Gamma_{I+,j}(t) > 0$.

On the interval $[s_1, \sigma)$, we may define $\theta^{1,*} = e_{(I+,0)}$ and $\theta^{2,*} = e_{(0,J+)}$. Throwing balls into such urns has no effect on the occupancy distribution, meaning that $\Gamma^*(t) = \Gamma(\sigma)$ is the solution on this interval to the equation

$$\dot{\Gamma}^* = \sum_{k=1}^{2} \dot{x}_k T^k[\theta^{k,*}] = 0.$$

The desired property that $\theta_{i,j}(t) > 0$ implies $\Gamma_{i,j}(t) \geq \delta\sigma^k$ is trivially satisfied on this interval.

At time s_1, we have $x_1(s_1) = \beta_1^e(\Gamma^*(s_1))$ and $x_2(s_1) > \beta_2^e(\Gamma^*(s_1))$. On the interval $[s_2, s_1)$, we use a modified reverse-time induction in which

$$\tau_{m+1} = \max\{t : x_1(\tau_m) - x_1(t) = \Gamma_{u_1^m}^*(\tau_m) \text{ or } t = s_2\}.$$

On the m-th subinterval, the occupancy rates naturally are defined as $\theta^{1,*} = e_{u_1^m - (1,0)}$ and $\theta^{2,*} = e_{0,J+}$. Using similar arguments to before, it follows that

after a finite number of steps, the induction terminates with $\tau_M = s_2$, with $\Gamma^*_{0,J+}(s_2) \geq \Gamma_{0,J+}(\sigma)$, and with κ_1^m strictly decreasing in m. Since now $x_k(s_2) = \beta_k^e(\Gamma^*(s_2)), k = 1, 2$, the original induction argument may be used to continue the definition of Γ^* on $[0, s_2)$. Now both κ_1^m and κ_2^m are both strictly decreasing in m across the entire interval $[0, \sigma)$, and we have the desired property that $\theta_{i,j} > 0$ implies $\Gamma^*_{i,j}(s) \geq \Gamma_{i,j}(\sigma)$ for $s \in [0, \sigma]$.

This completes the construction of the processes $\Gamma^*, \theta^{1,*}, \theta^{2,*}$ on $[0, \sigma]$. Note that

$$\begin{aligned}
&\int_0^\sigma \left[\dot{x}_1 R(\theta^{1,*} \| \Gamma^*) + \dot{x}_2 R(\theta^{2,*} \| \Gamma^*)\right] ds \\
&\quad \leq \; -\int_0^\sigma \left[\dot{x}_1 \log\left(\delta\sigma^K\right) + \dot{x}_2 \log\left(\delta\sigma^K\right)\right] ds \\
&\quad = \; -\sigma \log\left(\delta\sigma^K\right) \\
&\quad \leq \; \varepsilon/2,
\end{aligned}$$

and that Γ^* deviates no more than ε from Γ on $[0, \sigma]$, while ending up at the same place: $\Gamma^*(\sigma) = \Gamma(\sigma)$.

The construction on $[\sigma, \tau]$ is simpler. Let $M \in \mathbb{N}$, and observe that $\Gamma_{i,j}(s)$ is uniformly bounded away from zero for all i, j and $s \in [\sigma, \tau]$. We partition $[\sigma, \tau]$ into M subintervals of length $c_M = (\tau - \sigma)/M$. On each interval we set

$$\theta^{k,*}_{i,j}(s) = \int_{\sigma + lc_M}^{\sigma + (l+1)c_M} \dot{x}_k(r)\theta^k_{i,j}(r) dr \Big/ \left[x_k(\sigma + (l+1)c_M) - x_k(\sigma + lc_M)\right]$$

if $\sigma + lc_M \leq s \leq \sigma + (l+1)c_M$ (the definition is unimportant if $x_k(\sigma + (l + 1)c_M) - x_k(\sigma + lc_M) = 0$). For $s \in [\sigma, \tau]$ let

$$\dot{\Gamma}^* = \dot{x}_1 T^1\left[\theta^{1,*}\right] + \dot{x}_2 T^2\left[\theta^{2,*}\right], \Gamma^*(\sigma) = \Gamma(\sigma).$$

Since $\Gamma_{i,j}(s)$ is uniformly bounded away from zero, it is easy to check that for large enough M, Γ^* is a valid occupancy path that is associated with the processes $\theta^{1,*}, \theta^{2,*}$ and x. The convergence $\Gamma^* \to \Gamma$ on $[\sigma, \tau]$ as $M \to \infty$ is immediate, and it follows from the Lebesgue Dominated Convergence Theorem that

$$\begin{aligned}
&\lim_{M\to\infty} \int_\sigma^\tau \left[\dot{x}_1 R(\theta^{1,*} \| \Gamma^*) + \dot{x}_2 R(\theta^{2,*} \| \Gamma^*)\right] ds \\
&\quad = \int_\sigma^\tau \left[\dot{x}_1 R(\theta^1 \| \Gamma) + \dot{x}_2 R(\theta^2 \| \Gamma)\right] ds.
\end{aligned}$$

Therefore all parts of the lemma hold for sufficiently large M. □

5 Proof of the Large Deviation Lower Bound

Theorem 5.1 *Under Assumption 4.1*

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\exp\left[-nF(\Gamma^n)\right] \le \inf_{\Gamma\in\mathcal{S}}\left[I(\Gamma)+F(\Gamma)\right].$$

Proof: Consider any Γ for which $I(\Gamma)<\infty$. Then it suffices to show that

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\exp\left[-nF(\Gamma^n)\right] \le I(\Gamma)+F(\Gamma).$$

Owing to Lemma 4.5 and the continuity of F, we can assume without loss that there are $\delta>0$ and $K\in\mathbb{N}$ such that $\Gamma_{i,j}(t)\ge\delta t^K$.

We again utilize the representation (2.3). Fix $b>0$. According to the representation, the inequality in the last display will follow if we can find a sequence $\{\mu^n, n\in\mathbb{N}\}$ such that

$$\limsup_{n\to\infty}\frac{1}{n}R\left(\mu^n\|\Pi^n\otimes\Lambda^n\right)\le I(\Gamma)+b, \tag{5.1}$$

and such that if $\bar{\Gamma}^n$ is the urn process constructed under the distribution μ^n, then

$$\limsup_{n\to\infty}\bar{P}^n\left\{d(\bar{\Gamma}^n,\Gamma)>b\right\}\le b. \tag{5.2}$$

To prove the desired bound we must construct an appropriate sequence of measures μ^n. For Γ as above, let θ^1,θ^2 and $\tilde{x}$ denote corresponding rate and coloration processes which achieve the infimum. Without loss we can assume these processes satisfy the properties ascribed to $\theta^{1,*}$ and $\theta^{2,*}$ in Lemma 4.7.

For $a>0$ let $G:\mathcal{T}\to\mathbb{R}$ be continuous and satisfy

$$G(x)=\begin{cases}1/a & \text{if } d(x,\tilde{x})\ge 2a\\ 0 & \text{if } d(x,\tilde{x})\le a,\end{cases}$$

and also $G(x)\in[0,1/a]$ for all $x\in\mathcal{T}$. Since $\{x^n, n\in\mathbb{N}\}$ satisfies a large deviation principle with rate function J, there is a sequence $\{\lambda^n, n\in\mathbb{N}\}$, satisfying

$$\begin{aligned}\bar{E}^n\left[\frac{1}{n}R\left(\lambda^n\|\Lambda^n\right)+G(\bar{x}^n)\right] &\to \inf_{x\in\mathcal{T}}\left[J(x)+G(x)\right]\\ &\le J(\tilde{x}).\end{aligned}$$

We use these measures in constructing μ^n by setting

$$
\begin{aligned}
&\mu^n(dx_1, \ldots, dx_{N_n}, dy_1, \ldots, dy_{N_n}) \\
&\quad = \mu_1^n(dx_1|y_1, \ldots, y_{N_n}) \cdots \mu_{N_n}^n\ (dx_{N_n}|x_1, \ldots, x_{N_n-1}, y_1, \ldots, y_{N_n}) \\
&\qquad \lambda^n(dy_1, \ldots, dy_{N_n}).
\end{aligned}
$$

We will need to know how much λ^n mass is placed on sequences $y_1, \ldots, y_{\lfloor n\tau \rfloor + 1}$ such that if $\bar{x}^n$ is the corresponding cumulative coloration process, then

$$
\sup_{t\in[0,\tau]} d(\bar{x}^n(t), \tilde{x}(t)) \geq 2a.
$$

We have

$$
\begin{aligned}
\limsup_{n\to\infty} \frac{1}{a} \bar{P}^n \left\{ \sup_{t\in[0,\tau]} d(\bar{x}^n(t), \tilde{x}(t)) \geq 2a \right\} &\leq \limsup_{n\to\infty} \bar{E}^n \left[G(\bar{x}^n) \right] \\
&\leq \inf_{x\in\mathcal{T}} \left[J(x) + G(x) \right] \\
&\leq J(\tilde{x}) \\
&< \infty.
\end{aligned}
$$

Thus the probability $\bar{P}^n \left\{ \sup_{t\in[0,\tau]} d(\bar{x}^n(t), \tilde{x}(t)) \geq 2a \right\}$ can be made as small as desired for large n by taking a small. Since $\|F\|_\infty < \infty$, when constructing the controlled urn process we will be able to ignore these paths, and can let the measures that select the urns be the original uniform measure for such points in the underlying probability space. The relative entropy cost for such paths is then zero. Thus in the rest of this construction we focus on the case where the underlying coloration process satisfies

$$
\sup_{t\in[0,\tau]} d(\bar{x}^n(t), \tilde{x}(t)) \leq 2a.
$$

To finish the construction we must specify the conditional distributions of the X_l^n. Note that when specifying these distributions we get to see the complete outcome $Y_1^n, \ldots, Y_{N_n}^n$ of the coloration process, and can assume that this process satisfies the last equation.

The construction naturally separates according to the partition $[0, \tau] = [0, \sigma) \cup [\sigma, \tau]$. We must specify the distribution of the measures μ_l^n (or equivalently, the distribution of the random measures $\bar{\mu}_l^n$) for $l = 1, \ldots, N_n$. Recall that in general these measures are allowed to depend on the "past" $X_q^n, q < l$ However, it will turn out that we can assign μ_l^n based on just the coloration sequence and the time index l (i.e., "open loop" controls). Let (s_1^m, s_2^m) be

denote the finite collection of intervals on which $\theta^k_{i,j}(t)$ is constant, so that these intervals are nonoverlapping, and $[0,\tau]\setminus\cup_{m=1}^M (s_1^m,s_2^m)$ consists of a finite number of points. Suppose $l/n\in[s_1^m,s_2^m)$.

- If $s_2^m\leq\sigma$ then for each k there is (i,j) such that $\theta^k_{i,j}(t)=1$ for $t\in(s_1^m,s_2^m)$. If $Y_l^n=k$ (the ball at time l is color k), then μ_l^n set to be the uniform distribution on all urns of class (i,j). Note that when we rewrite the relative entropy as in (3.2) there will be equality, and in fact

$$\begin{aligned}\bar{E}^n\left[R\left(\bar{\mu}_l^n\|\pi^n\right)\right] &= \bar{E}^n\left[R\left(\nu^n(l/n)\|\bar{\Gamma}^n(l/n)\right)\right]\\ &= \bar{E}^n\left[\sum_{i=0,j=0}^{I+,J+}\left[\log\left(\frac{\nu^n_{i,j}(l/n)}{\bar{\Gamma}^n_{i,j}(l/n)}\right)\ \nu^n_{i,j}(l/n)\right]\right]\\ &= -\bar{E}^n\left[\log\left(\bar{\Gamma}^n_{i,j}(l/n)\right)\right].\end{aligned}$$

- If $s_2^m>\sigma$ then the controls are no longer "pure." If $y_l=k$ then each $\theta^k_{i,j}(t)$ determines a "weight" that should be placed on urns of class (i,j). We let μ_l^n be the measure which places mass $\theta^k_{i,j}(t)$ on the urns of class (i,j), and within this class uses the uniform distribution to apportion mass. We again have equality in the relative entropy in (3.2), and in fact

$$\begin{aligned}\bar{E}^n\left[R\left(\bar{\mu}_l^n\|\pi^n\right)\right] &= \bar{E}^n\left[R\left(\nu^n(l/n)\|\bar{\Gamma}^n(l/n)\right)\right]\\ &= \bar{E}^n\left[\sum_{i=0,j=0}^{I+,J+}\left[\log\left(\frac{\theta^k_{i,j}(l/n)}{\bar{\Gamma}^n_{i,j}(l/n)}\right)\ \theta^k_{i,j}(l/n)\right]\right].\end{aligned}$$

Note that for the controls constructed in Lemma 4.7 we have $\Gamma_{i,j}(t)\geq\delta\sigma^K$ for any (i,j) for which $\theta^1_{i,j}(t)\vee\theta^2_{i,j}(t)>0$. Owing to the randomness of the prelimit processes, we cannot guarantee the corresponding result $\bar{\Gamma}^n_{i,j}(t)\geq\delta\sigma^K$ for any (i,j) for which $\theta^1_{i,j}(t)\vee\theta^2_{i,j}(t)>0$. We therefore use a stopping time argument in the construction of the measures μ_l^n. Let $\bar{l}_n$ be the first time l such that $\bar{\Gamma}^n_{i,j}(l/n)\leq\delta\sigma^K/2$ for some (i,j) for which $\theta^1_{i,j}(l/n)\vee\theta^2_{i,j}(l/n)>0$. From time $\bar{l}_n$ on the construction above is modified, in that the measure is selected so that $\nu^n(l/n)=\bar{\Gamma}^n(l/n)$ for $l\geq\bar{l}_n$. Thus a weight of $\bar{\Gamma}^n_{i,j}(l/n)$ is placed on the urns of class (i,j). Note that with this definition $\bar{E}^n\left[R\left(\nu^n(l/n)\|\bar{\Gamma}^n(l/n)\right)\right]=0$ for $l\geq\bar{l}_n$. We now apply Theorem 3.1. Thus given any subsequence we have convergence along a further

subsequence as indicated in the theorem, with limit $(\bar{\Gamma}, \bar{x}_1, \bar{x}_2, \bar{\theta}^1, \bar{\theta}^2)$. Using the standard argument by contradiction, it will be enough to prove the convergence of controlled processes and bounds on the relative entropy cost for this convergent subsequence. Let $\gamma^n = (\bar{l}_n/n) \wedge \tau$. Note that because the applied controls are pure, the process $\bar{\Gamma}^n(t)$ is deterministic prior to σ, and also that prior to this time, the time derivative of both $\bar{\Gamma}^n(t)$ and $\Gamma(t)$ are piecewise constant. In fact, the two derivatives are identical except possibly on a bounded number of intervals of length less then $1/n$ (located near the endpoints of the intervals of constancy of $\dot{\Gamma}(t)$). Thus for large n we cannot have $\gamma^n < \sigma$. Since the range of γ^n is bounded we can also assume γ^n converges (along the same subsequence) in distribution to a limit γ, and it is easy to check that the limit control processes a.e. satisfy

$$\bar{\theta}^k_{i,j}(t) = \begin{cases} \theta^k_{i,j}(t) & \text{if } t \leq \gamma \\ \bar{\Gamma}_{i,j}(t) & \text{if } t > \gamma. \end{cases}$$

Owing to the definition of γ^n, if $\gamma < \tau$ then $\bar{\Gamma}_{i,j}(\gamma) = \delta\sigma^K/2$ for some (i,j). Recall also that $\Gamma_{i,j}(t) \geq \delta\sigma^K$ for all $t \in [\sigma, \tau]$.

Observe that the limit processes all implicitly depend on $a > 0$ through the function G. We claim that for each $b > 0$

$$\lim_{a \downarrow 0} \bar{P}\left\{ d(\bar{\Gamma}, \Gamma) > b \right\} = 0.$$

We already know that

$$\lim_{a \downarrow 0} \bar{P}\left\{ d(\bar{x}, \tilde{x}) > 2a \right\} = 0$$

However, since the rate processes $\theta^k_{i,j}$ are all piecewise constant, the integral

$$\int_0^t \left(\dot{\bar{x}}_1 T^1 \left[\theta^1 \right] + \dot{\bar{x}}_2 T^2 \left[\theta^2 \right] \right) ds$$

converges uniformly to

$$\int_0^t \left(\dot{\tilde{x}}_1 T^1 \left[\theta^1 \right] + \dot{\tilde{x}}_2 T^2 \left[\theta^2 \right] \right) ds$$

as $d(\bar{x}, \tilde{x}) \to 0$. Therefore,

$$\lim_{a \downarrow 0} \bar{P}\left\{ \sup_{0 \leq t \leq \gamma} \left\| \bar{\Gamma}(t) - \Gamma(t) \right\| > b \right\} = 0.$$

If $b > 0$ is sufficiently small, then following three items, all of which hold under $\gamma < \tau$, form a contradiction:

- $\bar{\Gamma}_{i,j}(\gamma) = \delta\sigma^K/2$ for some (i,j),
- $\Gamma_{i,j}(t) \geq \delta\sigma^K$ for all $t \in [0,\tau]$,
- $\sup_{0\leq t\leq\gamma} \left\|\bar{\Gamma}(t) - \Gamma(t)\right\| \leq b$.

We conclude that $\lim_{a\downarrow 0} \bar{P}\left\{\gamma < \tau\right\} = 0$, and therefore $\lim_{a\downarrow 0} \bar{P}\left\{d(\bar{\Gamma},\Gamma) > b\right\} = 0$ for all sufficiently small $b > 0$. It follows that given $b > 0$, for some fixed (sufficiently small) $a > 0$ $\limsup_{n\to\infty} \bar{P}\left\{d(\bar{\Gamma}^n,\Gamma) > b\right\} \leq b$.

We must also consider the relative entropy costs. However, again using the convergence $\lim_{a\downarrow 0} \bar{P}\left\{d(\bar{\Gamma},\Gamma) > b\right\} = 0$ and the dominated convergence theorem,

$$\begin{aligned}
&\limsup_{a\downarrow 0}\limsup_{n\to\infty} \bar{E}^n\left[\frac{1}{n}\sum_{l=1}^{N_n} R\left(\nu^n(l/n)\|\bar{\Gamma}^n(l/n)\right)\right] \\
&= \limsup_{a\downarrow 0}\limsup_{n\to\infty} \bar{E}^n\left[\sum_{m=1}^{M}\left(\sum_{l=1}^{N_n} R\left(\theta^1(l/n)\|\bar{\Gamma}^n(l/n)\right)\left(x_1^n\left((l+1)/n\right) - x_1^n\left(l/n\right)\right)\right)\right. \\
&\quad \left. + \left(\sum_{l=1}^{N_n} R\left(\theta^2(l/n)\|\bar{\Gamma}^n(l/n)\right)\left(x_2^n\left((l+1)/n\right) - x_2^n\left(l/n\right)\right)\right)\right] \\
&= \limsup_{a\downarrow 0} \bar{E}\left[\sum_{m=1}^{M}\int_{s_m^1}^{s_m^2} R\left(\theta^1(t)\|\bar{\Gamma}(t)\right)d\bar{x}_1(t) + \int_{s_m^1}^{s_m^2} R\left(\theta^2(t)\|\bar{\Gamma}(t)\right)d\bar{x}_2(t)\right] \\
&= \int_0^\tau \left[\dot{\tilde{x}}_1(t)R\left(\theta^1(t)\|\Gamma(t)\right) + \dot{\tilde{x}}_2(t)R\left(\theta^2(t)\|\Gamma(t)\right)\right]dt.
\end{aligned}$$

When combined with the bound

$$\limsup_{n\to\infty}\frac{1}{n}R\left(\lambda^n\left\|\Lambda^n\right.\right) \leq J(\tilde{x}),$$

for small enough $a > 0$ we have proved (5.1). Since $\lim_{a\downarrow 0} \bar{P}\left\{d(\bar{\Gamma},\Gamma) > b\right\} = 0$ implies (5.2) for small $a > 0$, the proof is complete. □

References

[1] S. Boucheron, F. Gamboa and C. Leonard. Bins and balls: Large deviations of the empirical occupancy process. *Annals of Applied Probability*, 2:1-30, 2002.

[2] P. Dupuis, R.S. Ellis, and A. Weiss. Large deviations for Markov processes with discontinuous statistics, I: General upper bounds. *Annals of Probability*, 19:1280–1297, 1991.

[3] P. Dupuis and R. Ellis. *A Weak Convergence Approach to Large Deviations.* Wiley Interscience, New York, 1997.

[4] P. Dupuis, C. Nuzman, and P. Whiting. Large deviations for occupancy problems. *Annals of Probability*, (to appear).

[5] P. Dupuis, C. Nuzman, and P. Whiting. Large devation asymptotics for occupancy systems with colored balls: The extremals, rate function and applications. *In preparation.*

[6] V. Eramo, M. Listanti, C. Nuzman, and P. Whiting, Optical switch dimensioning and the classical occupancy problem. *International Journal of Communications*, 15, pp. 127-141, March 2002.

[7] N.L. Johnson and S. Kotz. *Urn Models and their Applications.* John Wiley, 1977.

[8] S. Kotz and N. Balakrishnan. *Advances in Urn Models During the Past Two Decades.* In *Advances in Combinatorial Methods and Applications to Probability and Statistics, N. Balakrishnan Ed.*, pages 203–257, 1997.